

Strong Consistency of the Fréchet Sample Mean in Bounded Metric Spaces

Cedric E. Ginestet*

*Department of Mathematics and Statistics
Boston University, Boston, MA.*

Abstract: The Fréchet mean or barycenter generalizes the idea of averaging in spaces where pairwise addition is not well-defined. In general metric spaces, however, the Fréchet sample mean is not a consistent estimator of the theoretical Fréchet mean. For non-trivial examples, the Fréchet sample mean may fail to converge. Hence, it becomes necessary to consider other types of convergence. We show that a specific type of almost sure (a.s.) convergence for the Fréchet sample mean introduced by Ziezold (1977) is, in fact, equivalent to the consideration of the Kuratowski outer limit of a sequence of Fréchet sample means. Equipped with this outer limit, we prove different laws of large numbers for random variables taking values in a separable (pseudo-)metric space with a bounded metric. In this setting, we describe strong laws of large numbers for both the restricted and unrestricted Fréchet sample means of all orders, thereby generalizing Ziezold's original result. In addition, we also show that both the restricted and unrestricted Fréchet sample means are *metric* squared error (MSE) consistent. Interestingly, we derive a simple upper bound for this MSE, which is composed of the Fréchet variance of the estimator and a bias term, thereby generalizing the classical decomposition of the mean squared error for estimators of real-valued random variables.

AMS 2000 subject classifications: Barycenter, Centroid, Consistency, Estimation theory, Equicontinuity, Fréchet mean, Fréchet variance, Karcher Mean, Metric space, Metric squared error, Point function.

1. Introduction

All statistics are summaries. The epitome of these summaries is the sample mean, and its theoretical analog, the expected value. In an inspired monograph, Fréchet (1948) generalized this concept to any abstract metric space. He showed that the sole requirement for the definition of a mean element is the specification of a metric on the space of interest. Once this metric has been chosen and a probability measure has been defined on that metric space, the Fréchet mean is simply the element that minimizes the sum of the squared distances from all the

*This work was supported by a grant from the Air Force Office for Scientific Research (AFOSR), whose grant number is FA9550-12-1-0102; and by a fellowship from the UK National Institute for Health Research (NIHR) Biomedical Research Center for Mental Health (BRC-MH) at the South London and Maudsley NHS Foundation Trust and King's College London. Some portions of this work was conducted during a visit of the author at the University of Warwick, which was generously supported by the Center for Research in Statistical Methodology (CRiSM). We would also like to thank Pietro Rigo, Eric D. Kolaczyk, Thomas Nichols and Wilfrid S. Kendall for useful advice.

elements in that space. The Fréchet mean generalizes other notions of means in abstract spaces, such as the centroid in Euclidean geometry, the barycenter or center of mass in physics, the Procrustean mean in shape spaces (Le, 1998), and the Karcher mean on Riemannian manifolds (Karcher, 1977). The sample version of the Fréchet mean can naturally be expressed using cumulative addition instead of the expectation, thereby producing a convex combination operator on metric spaces of both negative and positive Alexandrov curvature (Ginestet et al., 2012).

The object of this paper is to characterize the asymptotic behavior of the Fréchet sample mean in separable bounded metric spaces, where a bounded metric space is a metric space with a bounded metric. Separability is a relatively mild topological assumption likely to be satisfied in most applications. The boundedness of the metric, however, is a more stringent condition. Nonetheless, there is a range of modern statistical applications for which the metric of interest is likely to be bounded. In bioinformatics, the use of the Hamming (1950) distance on finite alphabets, such as stretches of DNA for instance, naturally gives rise to such assumptions (He et al., 2004). Similarly, the comparison of families of networks with a given number of nodes, as commonly done in neuroscience (Ginestet et al., 2011) may similarly generate bounded metric spaces; albeit the combinatorial nature of these metrics may lead to bounds that increase factorially with the number of nodes in these networks. The Fréchet mean has proved to be especially popular in machine learning, where it has been applied to spaces of probability measures to facilitate clustering (Lee et al., 2007) and to spaces of images (Davis and Lazebnik, 2008, Gerber et al., 2009).

The asymptotic properties of the Fréchet sample mean have been studied by several authors. Ziezold (1977) proved a strong law of large numbers for Fréchet sample means defined in separable bounded quasi-metric spaces, where the metric is not assumed to satisfy the coincidence axiom. This a.s. convergence result has also been demonstrated for compact metric spaces by Sverdrup-Thygeson (1981). The perspectives adopted by these two authors are very different in nature. Given the fact that Sverdrup-Thygeson (1981) does not cite the work of Ziezold (1977), and because the work of the latter was published in a conference proceedings, it is probable that Sverdrup-Thygeson (1981) was not cognisant of Ziezold’s proof technique.

The result due to Ziezold (1977) is stronger than the one due to Sverdrup-Thygeson (1981). By the Heine-Borel theorem, a metric space is compact if, and only if, it is complete and totally bounded. The latter condition implies that every compact metric space has finite diameter, and therefore constitutes a bounded metric space. (Alternatively, using the continuity of the metric function, observe that the continuous mapping of a compact space is itself compact.) The converse, however, does not hold. A bounded metric space need not be compact: One can transform any metric space into a bounded metric space, by adopting the *discrete metric* (i.e. $d(x, y) = 1$ if x and y are identical and 0, otherwise). In general, an infinite set endowed with the discrete metric will be bounded, but not totally bounded, in the sense that it may not be possible to cover such a space with a finite number of balls of finite diameter.

The properties of sample Fréchet means on Riemannian manifolds have been particularly well-studied (Bhattacharya and Patrangenaru, 2002, 2005, Bhattacharya and Bhattacharya, 2012). When the Fréchet mean is assumed to be unique, the theorem of Sverdrup-Thygeson (1981) has been generalized by Bhattacharya and Patrangenaru (2003) for *proper* metric spaces. Recall that a metric space is proper, if and only if every bounded closed subsets of that space is compact (Sahib, 1998, Yang, 2011). By the Hopf-Rinow theorem, every complete and connected Riemannian manifold is a proper metric space. Thus, Bhattacharya and Patrangenaru (2003) have weakened the compactness assumption made by Sverdrup-Thygeson (1981), and their strong law of large numbers apply to manifolds, under some very mild conditions. Recently, Kendall and Le (2011) have further generalized these results with a weak law of large numbers and a central limit theorem for sequences of Fréchet sample means based on non-iid random variables taking values on a Riemannian manifold.

Here, we position ourselves in the general setting of Ziezold (1977), where we are not assuming any existing smooth structure. We will consider sequences of random variables taking values in separable quasi-metric spaces with a bounded metric. We generalize the seminal result of Ziezold (1977) to Fréchet sample means of any orders, and to *restricted* Fréchet sample means. The restricted Fréchet sample mean is the most ‘typical’ quantity chosen from the available sampled values. The computation of the unrestricted Fréchet sample mean in arbitrary metric spaces can indeed prove to be arduous, since this necessarily requires a minimization over a complex space. The difficulties that arise when estimating the Fréchet mean in shape spaces, for instance, have received special attention (Dryden and Mardia, 1998, Kume and Le, 2000, Le, 2001, 2004). Estimation issues have also been addressed in spaces of covariance matrices, where a range of different metrics can be considered (Arsigny et al., 2007, Dryden et al., 2009, Yang et al., 2011). The use of the restricted Fréchet mean may therefore be useful in practice, as it greatly simplifies the minimization procedure.

Importantly, we also clarify previous results on the asymptotic consistency of the Fréchet sample mean, by showing that the modes of convergence studied by Ziezold (1977) and Sverdrup-Thygeson (1981) are, in fact, equivalent to the consideration of the Kuratowski outer limit of a sequence of Fréchet sample means. This straightforward reformulation directly leads to a proof of the convergence of the Fréchet sample mean in metric squared error (MSE) to its theoretical analogue. Of independent interest is the fact that this MSE can be bounded above by the sum of the Fréchet variance of the estimator of interest and a bias term, which therefore provides a generalization of the classical decomposition of the mean squared error for real-valued random variables.

One of the core difficulties with the consideration of the asymptotic properties of Fréchet sample means is that such functions can be *multivalued*. That is, when the Fréchet sample mean is not unique, we obtain a random variable that is a set-valued function, which takes values in the power set of \mathcal{X} , or more precisely in the Borel σ -algebra of \mathcal{X} . It then becomes necessary to consider the convergence of multivalued functions. To this end, we resort to the tools of set-valued analysis, as described by Aubin and Frankowska (2009). This difficulty leads us

to consider different ‘types’ of convergence, depending on whether we require the Fréchet sample mean to converge, or are simply interested in evaluating the asymptotic behavior of the outer limit of that sequence.

The main innovation in this paper is our formal set-valued perspective. Note that our approach differs from the one of Bhattacharya and Bhattacharya (2012), since we have allowed the metric spaces of interest to be non-compact, and not necessarily manifolds. In particular, we identify the key role played by the Kuratowski outer limit in the sequence of Fréchet sample means. This paper therefore constitutes an extension of the work of Ziezold (1977) and Sverdrup-Thygeson (1981) to Fréchet means of all orders, and to restricted Fréchet mean. Moreover, we have emphasized the importance of point functions and of the Glivenko-Cantelli lemma. By positioning ourselves in bounded metric spaces, we do not have the nearest-point property, and therefore there is no guarantee that the Fréchet mean sets are closed sets in these spaces of interest. Some care must therefore be taken when evaluating the convergence of such sequences of sets.

This paper is organized as follows. In section 2, we introduce and study different types of a.s. convergence for sequences of Fréchet sample means, and show through counterexamples that the Kuratowski outer limit is better suited for this purpose. In section 3, we generalize the strong law of Ziezold (1977) to Fréchet sample means of all orders. Section 4 is devoted to the description of the restricted versions of the Fréchet sample mean, and a generalization of a result due to Sverdrup-Thygeson (1981) to bounded metric spaces, for random variables with closed support. In section 5, we derive equivalent results in terms of MSE consistency.

2. Sequences of Fréchet Sample Means

2.1. Empirical and Theoretical Fréchet Means

A separable space \mathcal{X} is endowed with a metric $d : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$. This produces a metric space, (\mathcal{X}, d) , with elements x . Let a probability space be denoted by $(\Omega, \mathcal{F}, \mathbb{P})$, and define a random variable, X , on that space, which takes values in $(\mathcal{X}, \mathcal{B})$. Here, \mathcal{B} is the Borel σ -algebra generated by the topology, τ on \mathcal{X} , induced by d . The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is assumed to be *complete*, in the sense that every subset of every null set is measurable. This is particularly convenient for constructing product spaces based on Ω that remain well-behaved. In addition, we define $\mu(B) := (\mathbb{P} \circ X^{-1})(B)$, for every $B \in \mathcal{B}$. Naturally, X is here assumed to be $(\mathcal{F}, \mathcal{B})$ -measurable. Such a random variable will be termed an *abstract-valued* random variable, which will be contrasted with the more standard real-valued random variables.

In this setting, we compute the most ‘central’ element. This is the element that has the smallest expected distance to all other elements in \mathcal{X} . This approach

allows us to define the following moments (Fréchet, 1948),

$$\Theta^r := \operatorname{arginf}_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x), \quad \text{and} \quad \sigma^r := \inf_{x' \in \mathcal{X}} \int_{\mathcal{X}} d(x, x')^r d\mu(x), \quad (1)$$

for every $0 < r < \infty$, and where $\Theta^r \subseteq \mathcal{X}$. Observe that we are using the superscript r on the Fréchet variance as a simple marker of the order of the exponentiated metric. Thus, in general, it will not be true that $(\sigma^r)^{1/r}$ simplifies to σ^1 .

These are commonly referred to as the Fréchet mean and variance when $r = 2$. For other choices of r , we will refer to these different Fréchet moments as Fréchet moments of *order* r . Note that if the *infimum* of $\mathbb{E}[d(x, x')^r]$ exists, then it is unique. However, the *argument of the infimum* may not necessarily exist and may not be unique. If such an argument does not exist, then $\Theta^r = \emptyset$. When the minimizer is not unique, the ensemble of minimizers is sometimes referred to as the *Fréchet mean set*. In particular, observe that if Θ is not a singleton, $\sigma^2 = \mathbb{E}[d(X, \theta)^2]$ for any $\theta \in \Theta$, will not, in general, be equivalent to $\mathbb{E}[d(X, \Theta)^2]$, where the distance between an element x and a non-empty subset A of \mathcal{X} is defined as $d(x, A) := \inf\{d(x, y) : y \in A\}$, with $d(x, \emptyset) = \infty$. In this paper, Fréchet mean and Fréchet mean set will be used interchangeably. Observe that when \mathcal{X} is a Hilbert space, endowed with the inner product metric, then there exists a unique global minimizer and Θ is therefore a singleton.

Analogously, for a given sequence of abstract-valued random variables $X_i : \Omega \mapsto \mathcal{X}$, for every $i = 1, \dots, n$, one may define the following Fréchet sample moments of the r^{th} order

$$\hat{\Theta}_n^r := \operatorname{arginf}_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^n d(X_i, x')^r \quad \text{and} \quad \hat{\sigma}_n^r := \inf_{x' \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^n d(X_i, x')^r. \quad (2)$$

Observe that, even for the sample versions of the Fréchet moments, these infima need not be attained, and therefore these quantities may be empty for each n . When there is no ambiguity as to the order of $\hat{\Theta}_n^r$, we will simply refer to this quantity as $\hat{\Theta}_n$, and similarly for Θ . In the sequel, an element of Θ and an element of $\hat{\Theta}_n$ will be respectively denoted by θ and $\hat{\theta}_n$. Our interest will mainly lie in considering Fréchet moments of the second order, albeit some examples will also be studied where $r = 1$. In general metric spaces, the empirical and theoretical Fréchet means may not be closed. (Take, for instance, the space $\mathcal{X} := \{-2\} \cup (-1, 1) \cup \{2\}$, with $1/2$ point masses at -2 and 2 .) However, It is easy to see that the Fréchet mean and Fréchet sample mean are closed subsets of \mathcal{X} , if \mathcal{X} is Polish.

Lemma 1. *For any metric space (\mathcal{X}, d) , Θ^r and the $\hat{\Theta}_n^r$'s are closed.*

Proof. Clearly, if $\Theta^r = \emptyset$, then $\operatorname{cl}(\Theta^r) = \Theta^r$ and similarly for the $\hat{\Theta}_n^r$'s. Now, fix $r = 1$, and consider the Fréchet mean set $\Theta \subseteq \mathcal{X}$. Recall that the boundary of Θ is defined as $\partial(\Theta) := \{x \in \mathcal{X} : d(\Theta, x) = d(\Theta^C, x) = 0\}$, where $\Theta^C := \mathcal{X} \setminus \Theta$. We proceed by contradiction. Assume that $\theta_0 \in \partial(\Theta)$ and $\theta_0 \notin \Theta$, then it

follows that there exists $\theta \in \Theta$, such that by the triangle inequality, $d(\theta_0, X) \leq d(\theta_0, \theta) + d(\theta, X)$, for every $X \in \mathcal{X}$. Taking the expectation, this gives

$$\mathbb{E}[d(\theta_0, X)] \leq d(\theta_0, \theta) + \mathbb{E}[d(\theta, X)] = \inf_{x' \in \mathcal{X}} \mathbb{E}[d(X, x')],$$

since $d(\theta_0, \Theta) = 0$, and using the definition of Θ in equation (1). Thus, θ_0 is optimal with respect to the infimum over \mathcal{X} . However, we have assumed that $\theta_0 \notin \Theta$, which leads to a contradiction, and therefore $\partial(\Theta) \subseteq \Theta$.

Next, consider the case of $r > 1$. Through a classical result on metric spaces (see, for instance Fréchet, 1948, p.229), we have

$$\left(\mathbb{E}[d(\theta_0, X)^r]\right)^{1/r} \leq \left(\mathbb{E}[d(\theta_0, \theta)^r]\right)^{1/r} + \left(\mathbb{E}[d(\theta, X)^r]\right)^{1/r},$$

for every $r \geq 1$, and the result immediately follows, using the same argument. The proof is identical for the $\hat{\Theta}_n^r$'s. \square

2.2. Convergence of Fréchet Sample Mean Sets

In this section, we study and compare different modes of convergence for set-valued random variables. In particular, note that our chosen modes of convergence differ from the ones used by Bhattacharya and Bhattacharya (2012), since we are not here assuming the compactness of the underlying metric space \mathcal{X} . Moreover, the target Fréchet mean set is also allowed to be empty, thereby making it difficult to implement the methods of Bhattacharya and Bhattacharya (2012).

For the Fréchet sample mean and its theoretical analogue, a.s. convergence could be defined in (\mathcal{X}, d) using sequences of random sets as follows,

$$\mathbb{P} \left[\left\{ \omega \in \Omega : \hat{\Theta}_n(\omega) \rightarrow \Theta \right\} \right] = 1, \quad (3)$$

where observe that Θ is here treated as a fixed subset of \mathcal{X} . The event in equation (3) will have probability one if the sequence of random sets, denoted $\hat{\Theta}_n$, converges a.s. in a set-theoretical sense such that

$$\liminf_{n \rightarrow \infty} \hat{\Theta}_n(\omega) = \limsup_{n \rightarrow \infty} \hat{\Theta}_n(\omega) = \Theta, \quad (4)$$

for almost every $\omega \in \Omega$, and where $\liminf S_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} S_m$, and $\limsup S_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} S_m$ denote the standard inner and outer limits of a sequence of subsets of \mathcal{X} . For most purposes, however, this type of convergence is too strong. In fact, this criterion does not hold for Fréchet sample means defined with respect to general abstract-valued random variables. There are many non-trivial examples of sequences of Fréchet sample means that diverge. Consider the following example adapted from the three-dimensional case described by Sverdrup-Thygeson (1981).

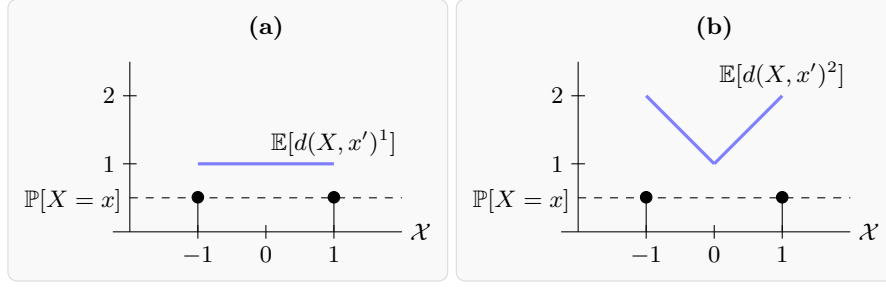


Fig 1. Metric and measure spaces considered in examples 1 and 2. In both panels, the closed interval $[-1, 1]$ is equipped with the Manhattan (or taxicab) metric, and two point masses are specified at -1 and 1 . Different Fréchet inferences are conducted by taking $r = 1$ and $r = 2$ in panels (a) and (b), respectively. In the first case, the theoretical Fréchet mean coincides with the median of X , whereas in panel (b), the theoretical Fréchet mean coincides with the arithmetic mean. However, the sequence of Fréchet *sample* means diverge in both cases, when convergence is evaluated using set-valued liminf and limsup, as described in equation (4).

Example 1. Let the interval, $\mathcal{X} := [-1, 1] \subset \mathbb{R}$, and equip this set with the usual ‘Manhattan’ distance, defined as $d(x, y) := |x - y|$ for every $x, y \in \mathcal{X}$. Additionally, let the random variable X , which takes values in \mathcal{X} , and which satisfies the following $\mathbb{P}[X = -1] = \mathbb{P}[X = 1] = 1/2$. This construction is illustrated in panel (a) of figure 1. The theoretical Fréchet mean of order $r = 1$ can be readily found as

$$\Theta^1 = \operatorname{arginf}_{x' \in \mathcal{X}} \sum_{x \in \{-1, 1\}} d(x, x') \mathbb{P}[x] = \mathcal{X},$$

since the energy function satisfies $\mathcal{E}(x') := \sum d(x, x') \mathbb{P}[x] = 1$ for every $x' \in \mathcal{X}$. Here, the Fréchet mean defined with respect to the Manhattan distance coincides with the *median* of the real-valued random variable X (Feldman and Tucker, 1966).

For the empirical Fréchet mean, $\hat{\Theta}_n^1$, first compute $S_n := \sum_{i=1}^n X_i$. Clearly, the S_n ’s are integer-valued. Observe the correspondence between the values of S_n and the values taken by the Fréchet sample mean. If the event $\{S_n = 0\}$ occurs, then it can easily be seen that $\hat{\Theta}_n$ is equal to \mathcal{X} . Similarly, $\{S_n \geq 1\}$, and $\{S_n \leq -1\}$ respectively imply that $\hat{\theta}_n = 1$ and $\hat{\theta}_n = -1$. Now,

$$\mathbb{P}[\{S_{2n} = 0\}] = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \approx (n\pi)^{-1/2},$$

for every n , using Stirling’s approximation. Since $\mathbb{P}[\{S_n = 0\}]$ is null, when n is odd, it follows that $\sum_{n=1}^{\infty} \mathbb{P}[\{S_n = 0\}] < \infty$, and therefore by the Borel-Cantelli lemma, we have $\mathbb{P}[\{S_n = 0\} \text{ i.o.}] = 0$, where i.o. means infinitely often. This implies that $\mathbb{P}[\{\hat{\Theta}_n = \mathcal{X}\} \text{ i.o.}] = 0$, and hence $\limsup \hat{\Theta}_n \neq \mathcal{X}$.

By using a similar argument, one can observe that $\mathbb{P}[\{S_n \leq -1\} \text{ i.o.}] = \mathbb{P}[\{S_n \geq 1\} \text{ i.o.}] = 1$, which implies that $\mathbb{P}[\{\hat{\theta}_n = -1\} \text{ i.o.}] = \mathbb{P}[\{\hat{\theta}_n = 1\} \text{ i.o.}] = 1$, and therefore $\{-1, 1\}$ is the limit superior of the sequence of Fréchet mean sets. By contrast, there does not exist an $N > 0$, such that $\hat{\theta}_n = 1$, for every $n \geq N$. An identical statement holds for $\hat{\theta}_n = -1$, and therefore the limit inferior of $\hat{\Theta}_n$ is empty. Thus,

$$\limsup_{n \rightarrow \infty} \hat{\Theta}_n(\omega) = \{-1, 1\} \supset \liminf_{n \rightarrow \infty} \hat{\Theta}_n(\omega) = \emptyset,$$

and the sequence of Fréchet sample means diverges, as criterion (4) is not satisfied.

Remark 1. The preceding example highlights two important aspects of the asymptotic behavior of the Fréchet sample mean set. Firstly, the Fréchet sample mean will in general fail to converge in the sense that its outer and inner limits need not be identical. In such cases, the sequence of Fréchet sample means exhibit an oscillatory property (see Feldman and Tucker, 1966). Secondly, the limit superior of a sequence of Fréchet sample means may solely represent a subset of the theoretical Fréchet mean. Taken together, these two problems necessitate (i) the study of the asymptotic behavior of the *outer limit* of the $\hat{\Theta}_n$'s, and (ii) the consideration of the convergence of the the Fréchet sample mean in terms of *set inclusion*, as a subset of the theoretical Fréchet mean. The passage from equations to inclusions is a natural step in the generalization of singleton-valued analysis to set-valued analysis.

Example 1 leads to the formulation of a weaker type of convergence, which can be expressed as the probability of the following event,

$$\left\{ \omega \in \Omega : \limsup_{n \rightarrow \infty} \hat{\Theta}_n(\omega) \subseteq \Theta \right\}. \quad (5)$$

However, we here encounter a slightly different problem than the one highlighted in our first example. This second issue can be illustrated through another counterexample, which shows that this particular type of a.s. convergence does not agree with the analogous real-valued a.s. convergence. That is, the reformulation of a given real-valued random variable into an abstract-valued setting, equipped with the same topology produces a divergent Fréchet sample mean in terms of equation (5). As a result, we obtain the somewhat counterintuitive result that the arithmetic sample mean differs from the corresponding Fréchet sample mean.

Example 2. Consider the same setting described in example 1, where now $r = 2$ (see panel (b) of figure 1). One can immediately see that the theoretical Fréchet mean is a singleton set,

$$\Theta^2 = \operatorname{arginf}_{x' \in \mathcal{X}} \sum_{x \in \{-1, 1\}} d(x, x')^2 \mathbb{P}[x] = 0,$$

which coincides with the expected value of the real-valued random variable X . For the Fréchet sample mean, we know from example 1 that $\mathbb{P}[\{S_n = 0\} \text{ i.o.}] = 0$ and therefore the probability of the sequence of empirical Fréchet means including $\mathbb{E}[X]$ infinitely often is null. That is, for $r = 2$, we have $\mathbb{P}[\{\hat{\theta}_n = 0\} \text{ i.o.}] = 0$. Observe that the same is true for any other specific sequence of realizations of X . Consider the case of $S_{3n} = nx_1 + 2nx_2$, where $x_1 = -1$ and $x_2 = 1$. For this subsequence, there exists a unique infimum, which is $\hat{\theta}_n = 1/3$. The probability of this event occurring is as follows,

$$\mathbb{P}[\{S_{3n} = nx_1 + 2nx_2\}] = \binom{3n}{n} \left(\frac{1}{2}\right)^{3n} \approx (1/2)^{5n},$$

which was approximated using Stirling's formulae. Clearly, all possible values of the Fréchet sample mean of X can be represented as a formulae of the form $nx_1 + \alpha nx_2$, for some $\alpha \in \mathbb{N}$. Using the Borel-Cantelli lemma, it therefore follows that there does not exist a point in $[-1, 1]$ that $\hat{\theta}_n$ will visit infinitely often, and hence $\limsup \hat{\Theta}_n = \liminf \hat{\Theta}_n = \emptyset$. By contrast, the arithmetic sample mean, $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$ trivially converges to the expected value of X a.s., since for every $\epsilon > 0$, there exists an $N > 1$, for which $d(\bar{X}_n(\omega), \mathbb{E}[X]) < \epsilon$, for every $n \geq N$, for almost every $\omega \in \Omega$. Thus, for this example, we reach the counterintuitive conclusion that $\bar{X}_n \notin \limsup \hat{\Theta}_n$, for every n .

This paradoxical disagreement between the divergence of the Fréchet sample mean and the classical convergence of the arithmetic sample mean in such a simple example requires a strengthening of our definition of the a.s. convergence of $\hat{\Theta}_n$. This particular problem seemed to have been implicitly identified by Ziezold (1977), as this author proposed the following type of convergence, which specializes the event presented in equation (5),

$$\left\{ \omega \in \Omega : \bigcap_{n=1}^{\infty} \overline{\bigcup_{m=n}^{\infty} \hat{\Theta}_m(\omega)} \subseteq \Theta \right\}, \quad (6)$$

where \bar{A} indicates the *closure* of set A in \mathcal{X} . For convenience, this particular type of convergence will be denoted by $\limsup \hat{\Theta}_n \subseteq \Theta$, a.s., where the limsup operator is here defined with respect to set inclusion on the power set of \mathcal{X} . It is easy to see why definition (6) resolves the issue illustrated in example 2. By taking the closure of $\bigcup_{m=n}^{\infty} \hat{\Theta}_m$, we include all the elements for which there exists a sequence of $\hat{\theta}_n$'s converging to $\mathbb{E}[X]$, and therefore

$$\mathbb{E}[X] \in \overline{\bigcup_{m=n}^{\infty} \hat{\Theta}_m},$$

for every n , which implies that $\limsup \hat{\Theta}_n = \{\mathbb{E}[X]\}$, as desired, thereby ensuring complete agreement between the classical and Fréchet inferential approaches for this particular example. Note that these issues are neither related to the

completeness of the underlying space of interest, nor associated to the question of the non-emptiness of Θ .

Since Sverdrup-Thygeson (1981) assumed that \mathcal{X} is compact, it follows that Θ and $\widehat{\Theta}_n$ are non-empty, in this case. The separability of \mathcal{X} is not sufficient to ensure that Θ and the $\widehat{\Theta}_n$'s are non-empty. Nonetheless, observe that if $\widehat{\Theta}_n = \emptyset$, then the events in equations (5) and (6) are trivially almost certain, since $\emptyset \subseteq A$, for all $A \subseteq \mathcal{X}$, as originally observed by Ziezold (1977).

2.3. Kuratowski Upper Limit

It can easily be shown that the type of convergence envisaged by Ziezold (1977) is, in fact, equivalent to the celebrated upper limit introduced by Kuratowski (1966), which has been adopted as the preferred type of convergence in set-valued analysis (see Aubin and Frankowska, 2009). The Kuratowski upper limit is defined over a metric space (\mathcal{X}, d) , for some sequence of subsets $A_n \subseteq \mathcal{X}$, as follows

$$\begin{aligned} \text{Limsup}_{n \rightarrow \infty} A_n &:= \left\{ x \in \mathcal{X} : \liminf_{n \rightarrow \infty} d(x, A_n) = 0 \right\} \\ &= \left\{ x \in \mathcal{X} : \{A_n \cap N_\epsilon(x) \neq \emptyset\} \text{ i.o., } \forall \epsilon > 0 \right\}, \end{aligned} \quad (7)$$

where \liminf and Limsup are taken with respect to real numbers and subsets of \mathcal{X} , respectively, and with $N_\epsilon(x) := \{x' \in \mathcal{X} : d(x, x') < \epsilon\}$. The second formulation of Limsup in equation (7) immediately follows from the positivity of the metric. Also, observe that the Kuratowski upper limit is equivalent to the set of *cluster points* of the sequences, $x_n \in A_n$ (Aubin and Frankowska, 2009). Clearly, the Kuratowski upper limit of any sequence of sets is *closed*, and moreover, it contains the conventional set-theoretical upper limit, such that for any sequence of random sets A_n ,

$$\limsup_{n \rightarrow \infty} A_n \subseteq \text{Limsup}_{n \rightarrow \infty} A_n.$$

Importantly, it can be easily shown that the Kuratowski upper limit and the quantity studied by Ziezold (1977) are equivalent, as stated in the following lemma.

Lemma 2. *Given a metric space (\mathcal{X}, d) , for any sequence of sets $A_n \subseteq \mathcal{X}$,*

$$\limsup_{n \rightarrow \infty} \overline{A_n} = \text{Limsup}_{n \rightarrow \infty} A_n.$$

Proof. Clearly, $\limsup_{n \rightarrow \infty} \overline{A_n} = \emptyset$, if and only if, $\text{Limsup}_{n \rightarrow \infty} A_n = \emptyset$. Thus, assume that these two outer limits are non-empty, and choose $x_0 \in \limsup_{n \rightarrow \infty} \overline{A_n}$. Then, $x_0 \in \bigcup_{m=N}^{\infty} \overline{A_m}$ for every N and there exists a subsequence x_k such that $x_k \in A_{n_k}$, for every k , which satisfies $x_k \rightarrow x_0$. Hence, we have $\liminf_{n \rightarrow \infty} d(x_0, A_n) = 0$, and by definition (7), $\limsup_{n \rightarrow \infty} \overline{A_n} \subseteq \text{Limsup}_{n \rightarrow \infty} A_n$.

Conversely, choose $x_0 \in \text{Limsup}_{n \rightarrow \infty} A_n$. Then, there exists a subsequence x_k such that $x_k \in A_{n_k} \cap N_\epsilon(x_0)$, for every k and for every $\epsilon > 0$, which satisfies $x_k \rightarrow x_0$,

as $k \rightarrow \infty$. Clearly, such subsequences can be found for every $N \in \mathbb{N}$, such that $n_1 \geq N$. This immediately implies that $x_0 \in \bigcap_{N=1}^{\infty} \bigcup_{m=N}^{\infty} A_m$, and therefore $\limsup A_n \supseteq \text{Limsup } A_n$, which completes the proof. \square

Observe that $\text{Limsup } A_n$ can be empty. Consider the following diverging sequence of sets, $A_n := [n-1, n+1]$, for every n . It is immediate that $\text{Limsup } A_n = \emptyset$. Throughout the rest of the paper, we will neither assume the existence nor the uniqueness of Θ^r and the Θ_n^r 's. In particular, in the sequel, Θ^r may be either empty, a set or a singleton set.

3. Almost Sure Consistency of Fréchet Sample Mean

In this section, we describe a generalization of the strong law of large numbers due to Ziezold (1977) to Fréchet sample means of any order. This generalization also allows us to re-formulate this original result using the Kuratowski upper limit.

Theorem 1. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a separable bounded metric space (\mathcal{X}, d) , let X_1, \dots, X_n be a sequence of independent and identically distributed (iid) abstract-valued random variables, such that $X_i : \Omega \mapsto \mathcal{X}$, for every X_i . Then,*

$$\widehat{\sigma}_n^r \rightarrow \sigma^r \quad a.s., \quad \text{and} \quad \text{Limsup}_{n \rightarrow \infty} \widehat{\Theta}_n^r \subseteq \Theta^r \quad a.s.,$$

for every finite $r \geq 1$, and where Limsup is defined as in equation (7).

The particular mode of convergence of the Fréchet sample mean used in theorem 1 will sometimes be denoted by $X_n \xrightarrow{a.s.} X$, which implies that $\text{Limsup } X_n \subseteq X$ with probability one.

Remark 2. The integrability of the r^{th} order metric is implied by the finiteness of both d and μ . Since $d(x, y) \leq M$, for every $x, y \in \mathcal{X}$, we have for any arbitrary $\alpha \in \mathcal{X}$ and finite $r \geq 1$,

$$\mathbb{E}[d(X, \alpha)^r] = \int_{\mathcal{X}} |d(x, \alpha)|^r d\mu(x) \leq \int_{\mathcal{X}} M^r d\mu(x) = M^r \mu(\mathcal{X}) < \infty,$$

by the linearity of the Lebesgue integral, and the fact that μ is a probability measure. Surprisingly, Ziezold (1977) originally assumed that $\mathbb{E}[d(X, \alpha)^r] < \infty$ for at least one $\alpha \in \mathcal{X}$. This condition, however, is redundant, since this author also assumed that the metric on \mathcal{X} is bounded and that the sample space of interest is a probability space. Although this only makes \mathcal{X} a bounded metric space, and not a totally bounded metric space (i.e. \mathcal{X} may not be coverable by a finite number of open balls with finite radii), the boundedness of d is nonetheless sufficient for deducing the integrability of d^r .

Remark 3. By contrast, the integrability of the exponentiated metric was not explicitly assumed by Sverdrup-Thygeson (1981). This author, however,

assumed that \mathcal{X} is compact, which implies that d^r is integrable for any finite $r \geq 1$. Indeed, compact metric spaces are totally bounded and therefore have bounded metric. If $d(x, y) \leq \text{diam}(\mathcal{X})$ for every $x, y \in \mathcal{X}$, then $d(x, y)^r \leq \text{diam}(\mathcal{X})^r < \infty$, and it therefore follows that $\int_{\mathcal{X}} |d(x, y)|^r d\mu(x) \leq \int_{\mathcal{X}} \text{diam}(\mathcal{X}) d\mu(x) = \text{diam}(\mathcal{X})\mu(\mathcal{X})$, for every $r \geq 1$, using again the linearity of the Lebesgue integral. Thus, if \mathcal{X} is compact, d is r -integrable with respect to any measure satisfying $\mu(\mathcal{X}) < \infty$.

The key to the proof of theorem 1 is based on a classical result, due to Rao (1962), which stipulates the conditions under which the weak convergence of a probability measure is equivalent to the uniform convergence of a probability measure, in a sense made clear in theorem 2. This can be seen as a generalization of the Glivenko-Cantelli lemma to random variables taking values in separable metric spaces (see also Parthasarathy, 1967, chap. 2). In this result, we will need to define a class of functions on the separable space \mathcal{X} , which we will denote by $\mathcal{F} := \mathcal{F}(\mathcal{X})$, whereby every $f \in \mathcal{F}$ is a real-valued continuous function that satisfies $f : \mathcal{X} \mapsto \mathbb{R}$. Such a class of functions is said to be *uniformly bounded* when for every $f \in \mathcal{F}$, and every $x \in \mathcal{X}$, there exists an $M \in \mathbb{R}$, such that $f(x) \leq M$. In addition, \mathcal{F} is *equicontinuous at a point* $x_0 \in \mathcal{X}$, if for every $\epsilon > 0$, there exists $\delta(x_0) > 0$, such that for every $u \in N_\delta(x_0) := \{u \in \mathcal{X} : d(x_0, u) < \delta\}$, we have $|f(x) - f(u)| < \epsilon$, for every $f \in \mathcal{F}$. The class \mathcal{F} is said to be equicontinuous if it is equicontinuous for every $x \in \mathcal{X}$. Finally, \mathcal{F} is said to be *uniformly equicontinuous* if δ does not depend on x_0 . We will denote the collection of all finite measures on \mathcal{B} by $\mathcal{M}(\mathcal{B})$, and \Rightarrow will indicate weak convergence.

Theorem 2 (Rao, 1962, p.672). *Let $\mathcal{F}(\mathcal{X})$ be a class of real-valued functions on a separable space \mathcal{X} , and assume that $\mathcal{F}(\mathcal{X})$ is (i) dominated by a continuous integrable function on \mathcal{X} , and that (ii) $\mathcal{F}(\mathcal{X})$ is equicontinuous. If, for some sequence of measures $\mu_n \in \mathcal{M}(\mathcal{B})$, and $\mu \in \mathcal{M}(\mathcal{B})$, we have $\mu_n \Rightarrow \mu$, a.s., then*

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \left| \int f d\mu_n - \int f d\mu \right| = 0, \quad \text{a.s..}$$

The following lemma will be used in the proof of theorem 1. This result links the properties of a bounded metric space with the conditions required in Rao's (1962) theorem. For this purpose, we will require the following classes of *point functions* on a metric space (see Searcoid, 2007).

Definition 1. *For any metric space (\mathcal{X}, d) , the z -point function is defined as $d_z(x) := d(z, x)$ for every $x \in \mathcal{X}$. The class of point functions on (\mathcal{X}, d) is then denoted by $\mathcal{D}(\mathcal{X}) := \{d_z : \forall z \in \mathcal{X}\}$. Similarly, we will make use of the class of exponentiated point functions, defined as follows,*

$$\mathcal{D}^r(\mathcal{X}) := \{d_z^r : \forall z \in \mathcal{X}\}.$$

for every finite $r \geq 1$, and where elements in either \mathcal{D} or \mathcal{D}^r will be denoted by d_z , and d_z^r , respectively, or simply z , when there is no ambiguity.

Lemma 3. *If (\mathcal{X}, d) is a bounded metric space, then $\mathcal{D}^r(\mathcal{X})$ is uniformly bounded and uniformly equicontinuous for every finite $r \geq 1$.*

Proof. By the boundedness of (\mathcal{X}, d) , there exists an $M \in \mathbb{R}$, such that $d(x, y) \leq M$, for every $x, y \in \mathcal{X}$. Therefore, $d_z(x) \leq M$, for every $x \in \mathcal{X}$, for every $d_z \in \mathcal{D}$, and thus \mathcal{D} is uniformly bounded. Moreover, since $d_z^r(x) \leq M^r < \infty$, for every finite $r \geq 1$, it follows that each \mathcal{D}^r also forms a uniformly bounded class of functions. Next, by the reverse triangle inequality, we have $|d_z(x) - d_z(x_0)| \leq d(x, x_0)$, for all $x, x_0, z \in \mathcal{X}$, thereby proving the (uniform) equicontinuity of the class \mathcal{D} on \mathcal{X} . For the case of $r \geq 1$, we consider the exponentiated version of the triangle inequality. Using the binomial expansion, we obtain

$$\begin{aligned} d(z, x)^r &\leq \left(d(z, x_0) + d(x_0, x) \right)^r \\ &= d(z, x_0)^r + \sum_{k=1}^{r-1} \binom{r}{k} d(z, x_0)^{r-k} d(x_0, x)^k + d(x_0, x)^r, \end{aligned}$$

and similarly, for any given $x_0 \in \mathcal{X}$, we have $d(z, x_0)^r \leq d(z, x)^r + \sum_{k=1}^{r-1} \binom{r}{k} d(z, x)^{r-k} d(x, x_0)^k + d(x, x_0)^r$. Combining these two inequalities and invoking the symmetry of d we have

$$\begin{aligned} |d(z, x)^r - d(z, x_0)^r| &\leq d(x_0, x)^r + d(x_0, x) M^{r-1} \sum_{k=1}^{r-1} \binom{r}{k} \\ &\leq d(x_0, x) M^{r-1} \left(1 + \sum_{k=1}^{r-1} \binom{r}{k} \right), \end{aligned}$$

where M is the uniform bound on the class \mathcal{D} . Now, choose $\delta = \epsilon / \gamma M^{r-1}$, where $\gamma := 1 + \sum_{k=1}^{r-1} \binom{r}{k}$, such that if $d(x, x_0) < \delta$, then $|d_z^r(x) - d_z^r(x_0)| < \gamma \delta M^{r-1} = \epsilon$, for every $x \in N_\delta(x_0)$, for every $d_z^r \in \mathcal{D}^r$, thence proving the equicontinuity of \mathcal{D}^r at x_0 . Since δ did not depend on the choice of x_0 , it follows that \mathcal{D}^r is also uniformly equicontinuous. \square

Proof of Theorem 1. Observe that the theorem is trivially verified if $\hat{\Theta}_n^r = \emptyset$. Thus, assume that $\hat{\Theta}_n^r$ is non-empty. We here adopt the line of argument followed by Sverdrup-Thygeson (1981). However, since we are not assuming compactness, there are several aspects of Sverdrup-Thygeson's proof that becomes somewhat delicate. In the sequel, we will make use of the following quantities formulated with respect to the class of point functions described in definition 1. For every $z \in \mathcal{X}$, let

$$T_n(z) := \frac{1}{n} \sum_{i=1}^n d_z^r(X_i) - \int_{\mathcal{X}} d_z^r(x) d\mu(x), \quad (8)$$

and similarly,

$$T_n^*(z) := \frac{1}{n} \sum_{i=1}^n d_z^r(X_i) - \int_{\mathcal{X}} d_{\theta}^r(x) d\mu(x). \quad (9)$$

Since $T_n(x)$ is real-valued, one can invoke the strong law of large numbers for real-valued random variables, which gives

$$T_n(z) \rightarrow 0, \quad \text{a.s.}, \quad \forall z \in \mathcal{X}. \quad (10)$$

Note, however, that since we have used infima in the definitions of the Fréchet theoretical and sample means in equations (1) and (2), it follows that the convergence of $T_n(z) \rightarrow 0$ is not assured when z is an element of Θ or an element of $\hat{\Theta}_n$. However, as established in lemma 3, the class of point functions, $\mathcal{D}^r(\mathcal{X})$, is uniformly bounded and (uniformly) equicontinuous. Moreover, in remark 2, we have seen that $\mathbb{E}[d_z^r(X)] < \infty$ is implied by the boundedness of d . Thus, it follows that there exists a continuous integrable function, i.e. $f(x) := M$, dominating every $d_z^r \in \mathcal{D}^r$. Moreover, a classical result on the convergence of empirical measures based on iid random variables taking values in separable metric spaces (see Parthasarathy, 1967, theorem 7.1, p.53) implies that

$$\mu_n \Rightarrow \mu, \quad \text{a.s.}, \quad (11)$$

where $\mu_n := n^{-1} \sum_{i=1}^n \delta_{X_i}$, is the empirical measure on \mathcal{X} . Therefore, we are in a position to apply theorem 2, which shows that the empirical measure, μ_n , converges uniformly with probability 1. That is,

$$\mathbb{P} \left[\sup_{z \in \mathcal{D}^r} \left| \frac{1}{n} \sum_{i=1}^n d_z^r(X_i) - \int_{\mathcal{X}} d_z^r(x) d\mu(x) \right| \rightarrow 0 \right] = 1,$$

which may be re-written as

$$\sup_{z \in \mathcal{D}^r} |T_n(z)| = \sup_{z \in \mathcal{X}} |T_n(z)| \rightarrow 0, \quad \text{a.s.} \quad (12)$$

Consequently, $T_n(\hat{\theta}_n) \rightarrow 0$, a.s., and $T_n(\theta) \rightarrow 0$, a.s., for every $\hat{\theta}_n \in \hat{\Theta}_n$ and every $\theta \in \Theta$, respectively.

Further, from the definition of $\hat{\theta}_n$ and θ , we can ‘sandwich’ $T_n^*(\hat{\theta}_n)$ in the following manner. Firstly, observe that by the minimality of the θ ’s,

$$\begin{aligned} T_n(\hat{\theta}_n) &= \frac{1}{n} \sum_{i=1}^n d_{\hat{\theta}_n}^r(X_i) - \int_{\mathcal{X}} d_{\hat{\theta}_n}^r(x) d\mu(x) \\ &\leq \frac{1}{n} \sum_{i=1}^n d_{\theta}^r(X_i) - \int_{\mathcal{X}} d_{\theta}^r(x) d\mu(x) = T_n^*(\hat{\theta}_n). \end{aligned} \quad (13)$$

Secondly, by the minimality of the $\hat{\theta}_n$ ’s, we similarly have,

$$\begin{aligned} T_n^*(\hat{\theta}_n) &= \frac{1}{n} \sum_{i=1}^n d_{\hat{\theta}_n}^r(X_i) - \int_{\mathcal{X}} d_{\hat{\theta}_n}^r(x) d\mu(x) \\ &\leq \frac{1}{n} \sum_{i=1}^n d_{\theta}^r(X_i) - \int_{\mathcal{X}} d_{\theta}^r(x) d\mu(x) = T_n(\theta). \end{aligned} \quad (14)$$

Thence, combining equations (13) and (14), we obtain,

$$T_n(\hat{\theta}_n) \leq T_n^*(\hat{\theta}_n) \leq T_n(\theta),$$

such that, using equation (12),

$$|T_n^*(\hat{\theta}_n)| \leq \max\{|T_n(\hat{\theta}_n)|, |T_n(\theta)|\} \rightarrow 0, \quad \text{a.s.}, \quad (15)$$

which proves the a.s. convergence of $\hat{\sigma}_n^r$ to σ^r .

We now turn to the convergence properties of the Fréchet sample mean of the r^{th} order, $\hat{\Theta}_n^r$. Here, we generalize Ziezold's (1977) proof strategy to Fréchet sample means of any order (see also Molchanov, 2005, p.185). Choosing

$$\hat{\theta} \in \text{Limsup}_{n \rightarrow \infty} \hat{\Theta}_n^r,$$

it then suffices to show that $\hat{\theta} \in \Theta^r$, which is verified if $\mathbb{E}[d(X, \hat{\theta})^r] \leq \mathbb{E}[d(X, x')^r]$, for every $x' \in \mathcal{X}$. We proceed by constructing the following subsequence of natural numbers.

Observe that from the definition of the Kuratowski upper limit and the equivalence relation reported in lemma 2, it follows that $\hat{\theta} \in \text{Cl}(\bigcup_{m=n}^{\infty} \hat{\Theta}_m^r)$, for every n , where $\text{Cl}(\cdot)$ denotes the closure of a set. Thus, one can construct a subsequence, $\{n_k : k \in \mathbb{N}\}$, such that for every k , there exists an element $\hat{\theta}_k \in \bigcup_{m=k}^{\infty} \hat{\Theta}_m^r$, which satisfies $d(\hat{\theta}_k, \hat{\theta}) \leq 1/k$. Moreover, we can define $n_k := \min\{n \in \mathbb{N} : n \geq k, \hat{\theta}_k \in \hat{\Theta}_n^r\}$. Now, from a standard consequence of Minkowski inequality, we have

$$\left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta})^r \right)^{1/r} \leq \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \right)^{1/r} + \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(\hat{\theta}_k, \hat{\theta})^r \right)^{1/r},$$

which gives

$$\left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta})^r \right)^{1/r} \leq \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \right)^{1/r} + \frac{1}{k}.$$

As $k \rightarrow \infty$, it then follows from equation (12) that since $(n_k)_{k \in \mathbb{N}}$ is a subsequence of $(n)_{n \in \mathbb{N}}$, we obtain

$$\left(\mathbb{E}[d(X, \hat{\theta})^r] \right)^{1/r} \leq \liminf_{k \rightarrow \infty} \left(\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \right)^{1/r}, \quad (16)$$

where \liminf is here taken with respect to non-negative real numbers. Moreover, by construction, each $\hat{\theta}_k$ is minimal with respect to any element $x' \in \mathcal{X}$, such that

$$\frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, \hat{\theta}_k)^r \leq \frac{1}{n_k} \sum_{i=1}^{n_k} d(X_i, x')^r, \quad (17)$$

for every $x' \in \mathcal{X}$ and $k \in \mathbb{N}$. Observe that given the continuity and monotonicity of $g(x) := x^{1/r}$ on the positive real numbers, we have $\liminf g(x_n) = g(\liminf x_n)$, for sequences satisfying $x_n \in \mathbb{R}^+$. Therefore, it suffices to combine equations (16) and (17) in order to obtain $\mathbb{E}[d(X, \hat{\theta})^r] \leq \mathbb{E}[d(X, x')^r]$, for every $x' \in \mathcal{X}$, as required. Thence, $\hat{\theta} \in \Theta^r$ a.s., but since $\hat{\theta}$ was arbitrary, we have $\text{Limsup } \hat{\Theta}_n^r \subseteq \Theta^r$ a.s., as required. \square

3.1. Extension to Equicontinuous Transformations

Our proof of theorem 1 crucially relies on the equicontinuity of the functional space $\mathcal{D}^r(\mathcal{X})$. One could therefore extend the previous results by allowing this class of functions to represent any continuous transformations of the point functions, $d_z(\cdot)$.

4. Restricted Fréchet Means

Theorem 1 can be extended to the case of the restricted Fréchet mean. This is a concept that was originally introduced and studied by Sverdrup-Thygeson (1981). Interest in restricted Fréchet means is motivated by the fact that the domain of some abstract-valued random variables may be too large to be optimized in a reasonable amount of time. In such cases, the Fréchet sample mean may be more suitably defined as one of the elements in the sample at hand. That is, consider the following definition of the *restricted* Fréchet sample mean and variance,

$$\hat{\Theta}_n^{*,r} := \operatorname{argmin}_{x' \in \mathbf{X}} \sum_{i=1}^n d(X_i, x')^r \quad \text{and} \quad \hat{\sigma}_n^{*,r} := \min_{x' \in \mathbf{X}} \sum_{i=1}^n d(X_i, x')^r,$$

where $\mathbf{X} := \{X_1, \dots, X_n\} \subseteq \mathcal{X}$ denotes the set of sampled variables. In practice, the sample mean is chosen among the available sampled iid realizations from X . In particular, observe that we employed the minimum instead of the infimum in the definitions of both $\hat{\Theta}_n^{*,r}$ and $\hat{\sigma}_n^{*,r}$, as the required optimal values necessarily exist, albeit they may not be unique. Hence, observe that $\hat{\Theta}_n^{*,r} \neq \emptyset$ for any n . Theoretical analogues of these restricted quantities can be defined as follows,

$$\Theta^{*,r} := \operatorname{argmin}_{x' \in W} \int_{\mathcal{X}} d(x, x')^r d\mu(x), \quad \text{and} \quad \sigma^{*,r} := \min_{x' \in W} \int_{\mathcal{X}} d(x, x')^r d\mu(x),$$

where W is the support of μ , denoted $\operatorname{supp}(\mu)$, which is assumed to be *closed*. Observe that this closure condition is required for ensuring that the Fréchet mean is contained in $\operatorname{supp}(\mu)$. For notational convenience, we will also assume in the sequel that $r = 2$, and omit that superscript. As previously, the elements of Θ^* and $\hat{\Theta}_n^*$ will be denoted by θ^* 's and $\hat{\theta}_n^*$'s, respectively. We here prove a generalization of a consistency result due to Sverdrup-Thygeson (1981) on the a.s. convergences of the restricted Fréchet sample mean and variance. Observe

that a restricted Fréchet sample mean can only converge to a restricted Fréchet mean.

Theorem 3. *Under the conditions of theorem 1, for every $r \geq 1$, and assuming that $\text{supp}(\mu)$ is closed,*

$$\widehat{\sigma}_n^{*,r} \rightarrow \sigma^{*,r} \quad \text{a.s.}, \quad \text{and} \quad \text{Limsup}_{n \rightarrow \infty} \widehat{\Theta}_n^{*,r} \subseteq \Theta^{*,r} \quad \text{a.s..}$$

Proof. Let us denote a quantity analogous to the ones defined in equations (8) and (9), but here based on the *restricted* theoretical Frechet mean,

$$\text{TR}_n^*(z) := \frac{1}{n} \sum_{i=1}^n d_z^r(X_i) - \int_{\mathcal{X}} d_{\theta^*}^r(x) d\mu(x), \quad (18)$$

where $\theta^* \in \Theta^*$. We will first demonstrate that

$$\min_{x' \in \mathbf{X}} \left| \text{TR}_n^*(x') - \text{TR}_n^*(\theta^*) \right| \rightarrow 0, \quad \text{a.s..} \quad (19)$$

In order to prove this a.s. convergence, we need the following quantity,

$$s(\delta) := \sup_{z \in W} \sup_{d(x,y) < \delta} \left| d_z^r(x) - d_z^r(y) \right|, \quad (20)$$

where the second supremum is taken over all pairs of elements $x, y \in W$, satisfying $d(x, y) < \delta$. Since the class of exponentiated point functions on \mathcal{X} , denoted \mathcal{D}^r , was shown to be uniformly equicontinuous in lemma 3, it follows that $s(\delta) \rightarrow 0$, as $\delta \rightarrow 0$. Moreover, it is straightforward to see that for every $\delta > 0$, we have

$$\begin{aligned} \sup_{d(x,y) < \delta} \left| \text{TR}_n^*(x) - \text{TR}_n^*(y) \right| &= \sup_{d(x,y) < \delta} \left| \frac{1}{n} \sum_{i=1}^n d_x^r(X_i) - \frac{1}{n} \sum_{i=1}^n d_y^r(X_i) \right| \\ &\leq \sup_{d(x,y) < \delta} \frac{1}{n} \sum_{i=1}^n \left| d_x^r(X_i) - d_y^r(X_i) \right| \\ &\leq s(\delta). \end{aligned}$$

Next, let $O_\delta := \{x \in \mathcal{X} : d(x, \theta^*) < \delta\}$, for any $\delta > 0$. Since $\theta^* \in \text{supp}(\mu)$, from the definition of the restricted Fréchet mean, it follows that $\mu(O_\delta) =: \alpha > 0$. Hence,

$$\mathbb{P}[\{X_1 \in O_\delta\} \cup \dots \cup \{X_n \in O_\delta\}] = 1 - \prod_{i=1}^n \mathbb{P}[\{X_i \notin O_\delta\}] = 1 - (1 - \alpha)^n,$$

which converges to 1, as $n \rightarrow \infty$, for any $\alpha > 0$. Moreover, observe that since $x' \in W$, for every $x' \in \mathbf{X}$, we also have

$$\limsup_{n \rightarrow \infty} \min_{x' \in \mathbf{X}} \left| \text{TR}_n^*(x') - \text{TR}_n^*(\theta^*) \right| \leq s(\delta).$$

It then suffices to let $\delta \rightarrow 0$, in order to obtain equation (19). Now, from the definitions of TR_n^* and T_n , it can be seen that $\text{TR}_n^*(\theta^*) = T_n(\theta^*)$, and therefore

$$\text{TR}_n^*(\hat{\theta}_n^*) = \min_{x' \in \mathbf{X}} \text{TR}_n^*(x') \leq T_n(\theta^*) + \min_{x' \in \mathbf{X}} |\text{TR}_n^*(x') - \text{TR}_n^*(\theta^*)|,$$

by the optimality of $\hat{\theta}_n^*$. This can be bounded below by using the minimality of θ^* , such that

$$\begin{aligned} T_n(\hat{\theta}_n^*) &= \frac{1}{n} \sum_{i=1}^n d_{\hat{\theta}_n^*}^r(X_i) - \int_{\mathcal{X}} d_{\hat{\theta}_n^*}^r(x) d\mu(x) \\ &\leq \frac{1}{n} \sum_{i=1}^n d_{\theta^*}^r(X_i) - \int_{\mathcal{X}} d_{\theta^*}^r(x) d\mu(x) = \text{TR}_n^*(\hat{\theta}_n^*). \end{aligned}$$

Combining the last two results, we obtain the following ‘sandwich’ inequality of $\text{TR}_n^*(\hat{\theta}_n^*)$,

$$T_n(\hat{\theta}_n^*) \leq \text{TR}_n^*(\hat{\theta}_n^*) \leq T_n(\theta^*) + \min_{x' \in \mathbf{X}} |\text{TR}_n^*(x') - \text{TR}_n^*(\theta^*)|.$$

Thence, this gives a.s.,

$$|\text{TR}_n^*(\hat{\theta}_n^*)| \leq \max \{ |T_n(\hat{\theta}_n^*)|, |T_n(\theta^*)| + \min_{x' \in \mathbf{X}} |\text{TR}_n^*(x') - \text{TR}_n^*(\theta^*)| \} \rightarrow 0,$$

using the strong law of large numbers on $T_n(\theta^*)$, and using equation (19) for the second term in the maximum. This proves that $\hat{\sigma}_n \rightarrow \sigma$, a.s.. The proof of $\text{Limsup } \hat{\Theta}_n^* \subseteq \Theta^*$ with probability 1, can be conducted using the same construction described in the proof of theorem 1, by choosing $\hat{\theta}^* \in \text{Limsup } \hat{\Theta}_n^*$, and noting that $\text{supp}(\mu)$ was assumed to be closed. \square

Remark 4. The use of uniform equicontinuity in the proof of theorem 3 requires special mention. Sverdrup-Thygeson (1981) was able to invoke the continuity of $s(\delta)$ with respect to δ in equation (20) by using the compactness of \mathcal{X} . Here, this property immediately follows from the uniform equicontinuity of the class of exponentiated point functions, $\mathcal{D}^r(\mathcal{X})$. This was the sole argument in the proof of Sverdrup-Thygeson (1981) for the a.s. convergence of the restricted Fréchet sample mean that required the compactness of \mathcal{X} . Hence, the boundedness of d constitutes a sufficient condition.

Remark 5. Under our assumptions and the ones postulated by both Ziezold (1977) and Sverdrup-Thygeson (1981), there is no guarantee that $\Theta \subseteq \text{supp}(X)$ holds, as assumed in the definition of the restricted Fréchet mean. In particular, one can easily construct a measure space where Θ belongs to a set of μ -measure zero. Consider the random variable described in example 2, where two point masses were located at -1 and 1 , respectively, and the Fréchet mean was computed with respect to the square of the ‘Manhattan’ distance. Clearly, the Fréchet mean is located in the barycenter of the interval $[-1, 1]$ but that center of mass does not belong to $\text{supp}(X)$.

One of the considerable advantages of using the restricted Fréchet sample mean is that its Kuratowski outer limit necessarily converges to a non-empty subset of \mathcal{X} . A generalization of the Bolzano-Weierstrass compactness theorem to set-valued analysis states that any sequence of subsets of a separable metric space converges to a (possibly empty) limit set (Aubin and Frankowska, 2009). In a similar fashion, we here make use of the classical Bolzano-Weierstrass theorem for real numbers in order to show that the Kuratowski outer limit of the restricted Fréchet sample mean in separable bounded metric spaces is non-empty with probability one. That is, we use the fact that the sequence of real-valued distances converges in order to deduce the almost sure non-emptiness of $\text{Limsup}_{n \rightarrow \infty} \hat{\Theta}_n^{*,r}$. Observe that a similar statement would not hold for the unrestricted Fréchet sample mean, as there is no guarantee that each element in the sequence of unrestricted Fréchet sample means is non-empty.

Theorem 4. *Under the conditions of theorem 1, and assuming that $\text{supp}(\mu)$ is closed, for every $r \geq 1$, we have $\Theta^{*,r} \neq \emptyset$ and*

$$\mathbb{P} \left[\left\{ \omega \in \Omega : \text{Limsup}_{n \rightarrow \infty} \hat{\Theta}_n^{*,r}(\omega) \neq \emptyset \right\} \right] = 1.$$

Proof. Since $\text{supp}(\mu)$ is closed it follows that both $\Theta^{*,r}$ and $\hat{\Theta}_n^{*,r}$'s are non-empty, as otherwise, there would exist a divergent sequence in $\text{supp}(\mu)$. For the limit superior of the sample Fréchet mean, choose an arbitrary reference point, $x_0 \in \mathcal{X}$. Clearly, for every $n \in \mathbb{N}$ and using the boundedness of the metric on \mathcal{X} , we have $d(\hat{\Theta}_n^{*,r}(\omega), x_0) \leq M < \infty$. By the Bolzano-Weierstrass theorem for sequences of bounded real numbers, it follows that for almost every $\omega \in \Omega$, there exists a subsequence n_k such that $d(\hat{\Theta}_{n_k}^{*,r}(\omega), x_0) \rightarrow C$, for some constant $C \in \mathbb{R}^+$.

Moreover, since the $\hat{\Theta}_n^{*,r}$'s are non-empty, it follows that one can choose a sequence $z_k \in \hat{\Theta}_{n_k}^{*,r}$, for every $k \in \mathbb{N}$, such that $z_k \rightarrow x_1$ with $d(x_1, x_0) = C$. This is always possible, since $d(\hat{\Theta}_{n_k}^{*,r}(\omega), x_0) \rightarrow C$. However, in that case, we obtain the following upper bound through an application of the triangle inequality,

$$d(\hat{\Theta}_{n_k}^{*,r}, x_1) \leq d(\hat{\Theta}_{n_k}^{*,r}, z_k) + d(z_k, x_1),$$

for every k . Taking the limit inferior, this gives,

$$\liminf_{k \rightarrow \infty} d(\hat{\Theta}_{n_k}^{*,r}, x_1) \leq \liminf_{k \rightarrow \infty} d(\hat{\Theta}_{n_k}^{*,r}, z_k) + \liminf_{k \rightarrow \infty} d(z_k, x_1),$$

where both terms on the right-hand side cancels out, since the z_k belongs to the Fréchet sample means and $z_k \rightarrow x_1$. Thus, $x_1 \in \text{Limsup}_{n \rightarrow \infty} \hat{\Theta}_n^{*,r}(\omega)$, and this holds for every $\omega \in \Omega$. \square

5. Metric Squared Error (MSE) Convergence

The convergence of an \mathcal{X} -valued random variable with respect to d^r is here denoted by

$$X_n \xrightarrow{d^r} X,$$

which signifies that

$$\lim_{n \rightarrow \infty} \mathbb{E} [d(X_n(\omega), X)^r] = 0.$$

Observe that we are here requiring that the limit of $\mathbb{E}[d(X_n(\omega), X)^r]$ is null, which is a stronger condition than what we have considered thus far, when evaluating the Kuratowski outer limit of the Fréchet sample mean in section 2. It will be shown, however, that such a stronger condition is satisfied, when both Θ^r and the $\hat{\Theta}_n^r$'s are non-empty.

5.1. Properties of d^r -Consistency

Equipped with this mode of convergence, we will be especially interested in considering the d^r -convergence of a sequence of empirical Fréchet means to the corresponding theoretical Fréchet mean. In this case, we will refer to this mode of convergence as the d^r -consistency of a random empirical Fréchet mean, with respect to a fixed theoretical Fréchet mean.

Definition 2. *The Fréchet sample mean, $\hat{\Theta}_n$, is said to be d^r -consistent, for some $r \geq 1$, with respect to the Fréchet mean, Θ , when $\hat{\Theta}_n \xrightarrow{d^r} \Theta$. The case of $r = 2$ will be referred to as metric squared error (MSE) convergence, and the MSE is defined as follows,*

$$\text{MSE}_d(\hat{\Theta}_n) := \mathbb{E}[d(\hat{\Theta}_n, \Theta)^2],$$

where the expectation is taken with respect to the n random variables in $\hat{\Theta}_n$.

The MSE is clearly reminiscent of the classical mean squared error for real-valued random variables. This analogy is strengthened by the fact that a straightforward upper bound can be derived for the MSE, which replicates the classical decomposition of the mean squared error for estimators of real-valued random variables. In order to derive this decomposition, it will be useful to introduce the concept of the Fréchet expectation of the sample estimator that we have considered so far. This expectation operator can be formally defined as follows, for any $r \geq 1$,

$$\mathbb{F}_d^r[\hat{\Theta}_n] := \operatorname{arginf}_{x' \in \mathcal{X}} \int \cdots \int_{\mathcal{X}^n} d(\hat{\Theta}_n(\mathbf{x}), x')^r d\mu^n(\mathbf{x}),$$

where μ^n is the complete product measure on \mathcal{X}^n and $\hat{\Theta}_n^r(\mathbf{x})$ is not necessarily the Fréchet sample mean, but could be any sample estimator, which is a function of \mathbf{x} . Using this notation, we can generalize the standard notion of the bias of a given estimator, as the supremum of the distance between the target parameter and its sample estimator. Thus, the d -bias is formally defined as follows

$$b_d^r(\hat{\Theta}_n) := \sup\{d(\hat{\theta}_n, \Theta)^r : \hat{\theta}_n \in \mathbb{F}_d[\hat{\Theta}_n]\}, \quad (21)$$

for any $r \geq 1$, where for notational convenience, we have assumed that both $\mathbb{E}[\cdot]$ and Fréchet means are taken with respect to $r = 2$. Similarly, the variance of an estimator can be constructed in the following manner,

$$\text{Var}_d^r(\hat{\Theta}_n) := \inf_{x' \in \mathcal{X}} \mathbb{E}[d(\hat{\Theta}_n(\mathbf{X}), x')^r],$$

where the expectation is taken with respect to the n -dimensional random vector $\mathbf{X} \in \mathcal{X}^n$. Equipped with these definitions, we can immediately derive a generalization of the classical decomposition of the mean squared error.

Lemma 4. *Given a sample estimator $\hat{\Theta}_n$ of Θ , we have*

$$\text{MSE}_d(\hat{\Theta}_n) \leq 2 \text{Var}_d(\hat{\Theta}_n) + 2b_d^2(\hat{\Theta}_n).$$

Proof. Since (\mathcal{X}, d) is a metric space, we can invoke the triangle inequality on $d(\hat{\Theta}_n, \Theta)$ with respect to an element $\hat{\theta}_n \in \mathbb{F}_d[\hat{\Theta}_n]$, which gives

$$d(\hat{\Theta}_n, \Theta)^2 \leq 2d(\hat{\Theta}_n, \hat{\theta}_n)^2 + 2d(\hat{\theta}_n, \Theta)^2,$$

It then suffices to take the expectation over \mathcal{X} , in order to obtain

$$\begin{aligned} \text{MSE}_d(\hat{\Theta}_n) &\leq 2\mathbb{E}[d(\hat{\Theta}_n, \hat{\theta}_n)^2] + 2d(\hat{\theta}_n, \Theta)^2 \\ &\leq 2 \text{Var}_d[\hat{\Theta}_n] + 2 \sup\{d(\hat{\theta}_n, \Theta)^2 : \hat{\theta}_n \in \mathbb{F}_d[\hat{\Theta}_n]\}, \end{aligned}$$

where the latter inequality follows from the fact that $\hat{\theta}_n$ is an argument minimizing $\mathbb{E}[d(\hat{\Theta}_n, x')^2]$, for the first term; and through an application of the definition of the d -bias in equation (21), for the second term. \square

Using the notation $\text{MSE}_d^r(\hat{\Theta}_n) := \mathbb{E}[d(\hat{\Theta}_n, \Theta)^r]$, lemma 4 can be readily generalized for any $r \geq 1$ to

$$\text{MSE}_d^r(\hat{\Theta}_n) \leq 2^{r-1} \left(\text{Var}_d^r[\hat{\Theta}_n] + b_d^r(\hat{\Theta}_n) \right),$$

using the generalized triangle inequality (see Fréchet, 1948, p.228). As for the standard convergence in r^{th} mean of real-valued random variables, convergence in d^r implies convergence in d^s when $r > s$, as described in the following lemma.

Lemma 5. *Given any sequence of \mathcal{X} -valued random variables, X_n and X , if $X_n \xrightarrow{d^r} X$, then $X_n \xrightarrow{d^s} X$, where $r > s \geq 1$.*

Proof. By the Lyapunov's inequality (Grimmett and Stirzaker, 2001), we have $\mathbb{E}[|Z|^s]^{1/s} \leq \mathbb{E}[|Z|^r]^{1/r}$, for any real-valued random variable Z and for every $r > s \geq 1$, and therefore

$$\left(\mathbb{E}[d(X_n(\omega), X)^s] \right)^{1/s} \leq \left(\mathbb{E}[d(X_n(\omega), X)^r] \right)^{1/r},$$

also holds for any abstract-valued random variables and every $r > s \geq 1$. The result follows by taking the limit with respect to n , on both sides, and noting that since X_n converges to X in d^r , it follows that $\mathbb{E}[d(X_n(\omega), X)^r]$ converges to 0 as $n \rightarrow \infty$. \square

5.2. MSE Consistency of Fréchet Sample Mean

The MSE consistency of the Fréchet sample mean in separable bounded metric spaces is strong provided that the target Fréchet mean and the elements of the sequence of empirical Fréchet means are non-empty. This holds for Fréchet means of all orders, i.e. for every $r \geq 1$, which can be shown to be d^s -consistent, for every $s \geq 1$. This theorem naturally follows from the previous results stating the a.s. convergence of this estimator, the properties of the Kuratowski outer limit and an application of the bounded convergence theorem.

Theorem 5. *Assume that the conditions of theorem 1 hold and that in addition, $\Theta^r, \hat{\Theta}_n^r \neq \emptyset$, for every n . Then, for every $r, s \geq 1$, it follows that we have $\hat{\Theta}_n^r \xrightarrow{d^s} \Theta^r$.*

Proof. By theorem 1, $\text{Limsup } \hat{\Theta}_n^r(\omega) \subseteq \Theta^r$, a.s., implies that every subsequence of $\hat{\Theta}_n^r(\omega)$ converges to a subset of Θ^r , and thus it immediately follows that

$$\lim_{n \rightarrow \infty} d\left(\hat{\Theta}_n^r(\omega), \Theta^r\right)^s = 0, \quad \text{a.s.},$$

for every finite $s \geq 1$, using $d(A, B) := \inf\{d(x, y) : x \in A, y \in B\}$, for every $A, B \subseteq \mathcal{X}$. In addition, observe that since (\mathcal{X}, d) is a bounded metric space, we have $d(\hat{\Theta}_n^r(\omega), \Theta^r) \leq M$, and therefore $d(\hat{\Theta}_n^r(\omega), \Theta^r)^s \leq M^s < \infty$, for every finite $r, s \geq 1$, $n \in \mathbb{N}$ and $\omega \in \Omega$. Thus, we can invoke the bounded convergence theorem, in order to take the expectation over the space Ω ,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[d(\hat{\Theta}_n^r(\omega), \Theta^r)^s \right] = 0,$$

for every finite $s \geq 1$, and this completes the proof. \square

Remark 6. A substantial advantage of this particular mode of convergence is that it automatically controls for the ‘emptiness’ of the Kuratowski outer limit of the sequence of Fréchet sample means. That is, if the outer limit of the sequence of empirical Fréchet mean sets is solely the empty set, then such a sample estimator will fail to be d^r -consistent. By contrast, the mode of convergence studied in section 3 would treat a sequence of sets with empty outer limit as a.s. consistent. This follows from the fact that $\emptyset \subseteq \Theta$, for any $\Theta \subseteq \mathcal{X}$. In such cases, a.s. consistency does not imply d^r -consistency.

Remark 7. Conversely, observe that for this particular mode of convergence, the Fréchet sample mean need not be entirely included in the theoretical mean. That is, the Fréchet sample and theoretical means solely need to have a non-empty intersection. One may encounter the situation where $\hat{\Theta}_n \cap \Theta \neq \emptyset$ and $\hat{\Theta}_n \not\subseteq \Theta$ occur, for infinitely many n . In this case, it follows that $d(\hat{\Theta}_n, \Theta) \rightarrow 0$, as $n \rightarrow \infty$, and therefore the Fréchet sample mean is d^1 -consistent, even though this does not imply a.s. consistency.

Thus, from remarks 6 and 7, it follows that a.s. convergence of the Limsup of the Fréchet sample mean and d^r -convergence do not in general imply each

other. We close this section with an immediate corollary ascertaining the d^r -convergence of the restricted Fréchet sample mean. The proof strategy used for theorem 5 can directly be employed in this setting. Here, however, we do not require to assume the non-emptiness of the restricted Fréchet mean or the elements of the sequence of empirical restricted Fréchet means, since the support of the underlying measure, μ , is already assumed to be closed.

Corollary 1. *Under the conditions of theorem 3, for every $r \geq 1$, $\hat{\Theta}_n^{*,r} \xrightarrow{d^s} \Theta^{*,r}$, for every finite $s \geq 1$.*

6. Conclusion

In this paper, we have generalized the results due to Sverdrup-Thygeson (1981) by relaxing the compactness assumption made by this author. This task has highlighted interesting links between the Sverdrup-Thygeson's proof and another classical proof of the a.s. convergence of the Fréchet sample mean, due to Ziezold (1977). In particular, we have shown that by assuming the boundedness of the metric of interest, we can deduce the uniform boundedness and uniform equicontinuity of any family of point functions on \mathcal{X} . These two properties were found to be required on two distinct occasions when proving asymptotic convergence results for the unrestricted and restricted Fréchet sample means, respectively. In the original proof of Sverdrup-Thygeson (1981), these two arguments rely on compactness, thereby showing that uniform boundedness and uniform equicontinuity constitute appropriate weaker assumptions.

Throughout, we have assumed that the underlying metric of interest is a full metric. However, as was originally done by Ziezold (1977), it can be shown that our results also hold for bounded pseudo-metrics, where one relaxes the *axiom of coincidence*. In this case, $d(x, y) = 0$ does not necessarily imply that $x = y$. It is easy to check that this particular property was not used in this paper, and therefore that all the aforementioned convergence theorems remain valid for Fréchet sample mean sets defined over separable bounded *pseudo-metric* spaces.

References

- Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, **29**(1), 328–347.
- Aubin, J.P. and Frankowska, H. (2009). *Set-Valued Analysis*. Birkhäuser, Boston.
- Bain, L. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics*. Duxbury, Pacific Grove, CA.
- Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds with Applications to Shape Spaces*. Cambridge University Press, New York.

- Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *The Annals of Statistics*, **33**(3), 1225–1259.
- Bhattacharya, R. and Patrangenaru, V. (2002). Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*, **108**(12), 23–35.
- Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics*, **31**(1), 1–29.
- Capiński, M. and Kopp, E. (2004). *Measure, Integral and Probability*. Springer-Verlag, New York.
- Davis, B. and Lazebnik, S. (2008). Analysis of human attractiveness using manifold kernel regression. In *15th IEEE International Conference on Image Processing*, 109–112.
- Dryden, I. and Mardia, K. (1998). *Statistical Analysis of Shape*. Wiley, London.
- Dryden, I., Koloydenko, A., and Zhou, D. (2009). Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*, **3**, 1102–1123.
- Feldman, D. and Tucker, H. (1966). Estimation of non-unique quantiles. *Annals of Mathematical Statistics*, **37**(2), 451–457.
- Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de L’Institut Henri Poincaré*, **10**(4), 215–310.
- Gerber, S., Tasdizen, T., Joshi, S., and Whitaker, R. (2009). On the manifold structure of the space of brain images. *Medical Image Computing and Computer-Assisted Intervention, MICCAI*, **5761**, 305–312.
- Ginestet, C., Nichols, T., Bullmore, E., and Simmons, A. (2011). Brain network analysis: Separating differences in cost from differences in topology. *PLoS ONE*, **6**(7), e21570.
- Ginestet, C., Simmons, A., and Kolaczyk, E. (2012). Weighted fréchet means as convex combinations in metric spaces: Properties and generalized median inequalities. *Probability and Statistics Letters*, **82**(10), 1859–1863.
- Grimmett, G. and Stirzaker, D. (2001). *Probability and random processes*. Oxford University Press, Oxford.
- Hamming, R. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, **29**, 147–160.
- He, M., Petoukhov, S., and Ricci, P. (2004). Genetic code, hamming distance and stochastic matrices. *Bulletin of Mathematical Biology*, **66**(5), 1405–1421.
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communication in Pure and Applied Mathematics*, **30**, 509–541.
- Kendall, W. and Le, H. (2011). Limit theorems for empirical fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, **25**(3), 323–352.
- Kume, A. and Le, H. (2000). Estimating fréchet means in bookstein’s shape space. *Advances in Applied Probability*, **32**(3), 663–674.
- Kuratowski, K. (1966). *Topology. Volume I. New Edition, Revised and Augmented*. Academic Press, New York.

- Le, H. (1998). On the consistency of procrustean mean shapes. *Advances in Applied Probability*, **30**(1), 53–63.
- Le, H. (2001). Locating fréchet means with application to shape spaces. *Advances in Applied Probability*, **33**(2), 324–338.
- Le, H. (2004). Estimation of riemannian barycentres. *LMS J. Comput. Math*, **7**, 193–200.
- Lee, S.M., Abbott, A., and Araman, P. (2007). Dimensionality reduction and clustering on statistical manifolds. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 1–7.
- Molchanov, I. (2005). *Theory of Random Sets*. Springer, London.
- Parthasarathy, K. (1967). *Probability Measures on Metric Spaces*. American Mathematical Society, London.
- Rao, R.R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, **33**(2), 659–680.
- Sahib, A. (1998). *Espérance d’une variable aléatoire à valeur dans un espace métrique*. Ph.D. thesis, Université de Rouen, France.
- Searcoid, M. (2007). *Metric spaces*. Springer, London.
- Sverdrup-Thygeson, H. (1981). Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. *The Annals of Statistics*, **9**(1), 141–145.
- Yang, L. (2011). *Médianes de mesures de probabilité dans les variétés Riemanniennes et applications à la détection de cibles radar*. Ph.D. thesis, Université de Poitiers, France.
- Yang, L., Arnaudon, M., and Barbaresco, F. (2011). Geometry of covariance matrices and computation of median. In *AIP Conference Proceedings*, vol. 1305, 479.
- Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians*.